



Heterogeneity in Cochrane and non-Cochrane meta-analyses in orthodontics

Koletsis, Despina ; Fleming, Padhraig S ; Michelaki, Iris ; Pandis, Nikolaos

Abstract: **OBJECTIVES** Heterogeneity describes the percentage of variability across the study effects that can be attributed to between-study differences in a meta-analysis. The aim of this project was to explore the magnitude of heterogeneity in Cochrane and non-Cochrane meta-analyses in orthodontic research and to identify possible associations between heterogeneity (I) and a number of study characteristics including number of studies, type of outcome and type of analysis. **METHODS** The contents of five major orthodontic journals and the Cochrane Database of Systematic Reviews were electronically searched from January 2000 to December 2017 to identify Systematic Reviews (SRs) with at least one meta-analysis. Included records were screened for reporting of I classified into four categories: 0%, 1-29%, 30-59%, 60-100%. Associations between I and review-level and synthesis-level characteristics were tested. Univariable and multivariable mixed effects ordinal logistic regression was used to identify significant predictors for statistical heterogeneity. **RESULTS** A total of 72 SRs comprising 391 meta-analyses were included with the majority based on non-Cochrane reviews (n = 54, 75%). Overall, 125 meta-analyses (32%) reported heterogeneity explained by chance (I = 0%), whereas high values of I (60-100%) were seen in 152 syntheses (39%). In the multivariable analysis, inclusion of each additional study within the synthesis presented 20% higher odds for substantial/considerable heterogeneity compared to lower heterogeneity categories (OR = 1.20; 95% CIs: 1.09, 1.31; p < 0.001). Use of fixed effect analysis (OR = 0.25; 95% CIs: 0.12, 0.55; p = 0.001) was associated with significantly lower odds. Cochrane versus non-Cochrane meta-analyses were not associated with higher odds for substantial/considerable heterogeneity (OR = 2.81; 95% CIs: 0.53, 14.91; p = 0.22). **CONCLUSIONS** Substantial statistical heterogeneity is present within a considerable number of orthodontic meta-analyses. Further efforts should be made to improve understanding of decisions to undertake meta-analyses and selection of studies eligible for inclusion. **CLINICAL SIGNIFICANCE** The consistency of meta-analyses could be improved with more careful consideration of individual study characteristics. Reduced heterogeneity in meta-analyses will ensue more solid evidence based decisions for clinical practice.

DOI: <https://doi.org/10.1016/j.jdent.2018.05.003>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-158373>

Journal Article

Accepted Version

Originally published at:

Koletsis, Despina; Fleming, Padhraig S; Michelaki, Iris; Pandis, Nikolaos (2018). Heterogeneity in Cochrane and non-Cochrane meta-analyses in orthodontics. *Journal of Dentistry*, 74:90-94.

DOI: <https://doi.org/10.1016/j.jdent.2018.05.003>

Title: Heterogeneity in Cochrane and non-Cochrane meta-analyses in orthodontics

Despina Koletsi^a, Padhraig S. Fleming^b, Iris Michelaki^c, Nikolaos Pandis^d

^a Clinic of Orthodontics and Paediatric Dentistry, Center of Dental Medicine, University of Zurich, Zurich, Switzerland and Private Practice in Athens, Greece

^b Department of Orthodontics, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

^c School of Dentistry, National and Kapodistrian University of Athens, Athens, Greece

^d Department of Orthodontics and Dentofacial Orthopedics, Dental School/Medical Faculty, University of Bern, Switzerland and Private Practice in Corfu, Greece

Short Title: Heterogeneity in orthodontic meta-analyses

Publication: Koletsi D, Fleming PS, Michelaki I, Pandis N. Heterogeneity in Cochrane and non-Cochrane meta-analyses in orthodontics. J Dent. 2018 May 5. pii: S0300-5712(18)30110-6. doi: 10.1016/j.jdent.2018.05.003. [Epub ahead of print]

Keywords: heterogeneity, I^2 , Cochrane, meta-analysis, orthodontic

Correspondence: Dr. Despina Koletsi, 5 Kanari St, 15127 Melissia, Attica, Greece; email: d.koletsi@gmail.com

Title: Heterogeneity in Cochrane and non-Cochrane meta-analyses in orthodontics

Abstract

Objectives: Heterogeneity describes the percentage of variability across the study effects that can be attributed to between-study differences in a meta-analysis. The aim of this project was to explore the magnitude of heterogeneity in Cochrane and non-Cochrane meta-analyses in orthodontic research and to identify possible associations between heterogeneity (I^2) and a number of study characteristics including number of studies, type of outcome and type of analysis.

Methods: The contents of five major orthodontic journals and the Cochrane Database of Systematic Reviews were electronically searched from January 2000 to December 2017 to identify Systematic Reviews (SRs) with at least one meta-analysis. Included records were screened for reporting of I^2 classified into four categories: 0%, 1- 29%, 30- 59%, 60- 100%. Associations between I^2 and review-level and synthesis-level characteristics were tested. Univariable and multivariable mixed effects ordinal logistic regression was used to identify significant predictors for statistical heterogeneity.

Results: A total of 72 SRs comprising 391 meta-analyses were included with the majority based on non-Cochrane reviews ($n=54$, 75%). Overall, 125 meta-analyses (32%) reported heterogeneity explained by chance ($I^2=0\%$), whereas high values of I^2 (~60-100%) were seen in 152 syntheses (39%). In the multivariable analysis, inclusion of each additional study within the synthesis presented 20% higher odds for substantial/considerable heterogeneity compared to lower heterogeneity categories (OR= 1.20; 95%CI: 1.09, 1.31; $p<0.001$). Use of fixed effect analysis (OR= 0.25; 95%CI: 0.12, 0.55; $p=0.001$) was associated with significantly lower odds. Cochrane versus non-Cochrane meta-analyses were not associated with higher odds for substantial/considerable heterogeneity (OR= 2.81; 95%CI: 0.53, 14.91; $p=0.22$).

Conclusions: Substantial statistical heterogeneity is present within a considerable number of orthodontic meta-analyses. Further efforts should be made to improve understanding of decisions to undertake meta-analyses and selection of studies eligible for inclusion.

Clinical Significance: The consistency of meta-analyses could be improved with more careful consideration of individual study characteristics. Reduced heterogeneity in meta-analyses will ensue more solid evidence based decisions for clinical practice

Introduction

Recent years has seen an increase in publication of systematic reviews (SRs) and meta-analyses in biomedical research as they are considered by healthcare professionals the most effective and transparent approach to provide a synthesis of research and clinical evidence and guide decision making in practice [1–3]. Notwithstanding this, these have attracted criticism both related to excessive publication of overlapping and barren reviews, as well as concerns in relation to the quality and yield associated with many of these [4–6].

A meta-analysis uses a mathematical approach to combine the results from individual studies with regard to a research question of interest in order to obtain an overall estimate of the treatment effect [7]. However, due to the fact that individual studies may differ in terms of settings, populations, interventions, outcomes or design, it may be possible that this lack of homogeneity or heterogeneity may render the findings inconsistent. Inconsistency refers to the presence of variation across the results of individual studies greater than that anticipated by chance [8]. Presence of unexplained inconsistency is an important reason for downgrading the quality of the evidence from meta-analyses [9,10].

Assessment of heterogeneity involves both clinical and statistical aspects. Clinical heterogeneity refers to *a priori* identification of differences in population settings, interventions or variations in treatment outcomes and may be detected upon examination of the eligibility criteria for the population under study [11,12]. Statistical heterogeneity can be inspected either visually or through statistical tests. Visual inspection involves examination of the forest plots and assessment of variation in the point estimates of the included studies as well as the degree of overlap of the associated confidence intervals [13]. Statistical heterogeneity may be confirmed through statistical tests, the most known of which are the Q- statistic also referred to as the chi-squared test and the

I^2 statistic [14,15]. A statistically significant chi-squared test (i.e. $p < 0.1$) may indicate important variation in individual study estimates; however, false negative results cannot be excluded especially when few studies are included in the meta-analysis [13]. The I^2 statistic is considered a more straightforward measure to quantify heterogeneity describing the percentage variability in the point estimates that are attributed to heterogeneity due to between study differences [14].

The Cochrane Collaboration has suggested the use of four categories when interpreting heterogeneity based on I^2 as follows: 0-40% unimportant heterogeneity, 30-60% moderate, 50-90% substantial and 75- 100% considerable heterogeneity. The degree of overlap between these categories represents the subjectivity in interpreting the magnitude of heterogeneity [16]. Empirical evidence based on Medline and Cochrane databases has indicated that moderate to substantial statistical heterogeneity has been identified in up to 39% of meta-analyses in medicine, depending on the type of outcome assessed, while I^2 has been shown to increase as the number of studies included in the meta-analysis increases for continuous outcomes [17].

The degree of statistical heterogeneity might constitute an indicator of inappropriate synthesis of conceptually different studies. Statistical heterogeneity may arise as a result of clinical or methodological differences between studies. This pertains to differences related to the participants, interventions and outcomes, or differences in study design or risk of bias within studies. Where the variation in treatment effect is greater than that anticipated to arise due to chance, the presence of increased levels of heterogeneity is suspected [10]. This in turn may compromise conclusions drawn from meta-analysis of heterogeneous studies, with attendant effects on the translation of the findings to clinical practice in terms of clinical decision making, or the generalizability of the results [18]. As such, careful consideration of study characteristics prior

to possible inclusion within a planned meta-analysis is important with clear delineation of eligibility criteria based ideally on a pre-planned protocol.

To our knowledge, there has been no similar attempt to quantify the magnitude of heterogeneity within meta-analyses in dentistry. The primary objective of the present study was therefore to record and quantify the extent of statistical heterogeneity (based on I-squared; I^2) in Cochrane and non- Cochrane meta-analyses in the major dental specialty of orthodontics. In addition, we aimed to identify possible associations between heterogeneity and a number of study characteristics including year of publication, funding, type and significance of outcome, type of mathematical analysis followed and number of studies included within the syntheses.

Materials and Methods

The contents of five major orthodontic journals, namely the American Journal of Orthodontics and Dentofacial Orthopedics (AJODO), Angle Orthodontist (Angle), European Journal of Orthodontics (EJO), Journal of Orthodontics (JO), and Orthodontics and Craniofacial Research (OCR), as well as orthodontic topics covered in the Cochrane Database of Systematic Reviews (CDSR) were electronically searched from January 2000 to December 2017. The selection of journals was based on the relative frequency of systematic reviews (SRs) published within them. On initial screening, articles were considered potentially eligible if they included the terms “systematic review” or “meta-analysis” in the title or the abstract. Relevant articles (SRs) were finally included after full text evaluation, when at least one mathematical synthesis was undertaken with associated report of the degree of heterogeneity based on I^2 statistic.

The selection of articles was carried out by two authors (DK, IM) and data was extracted on pre-specified standardized forms by one author (DK) covering the following information:

- Article level: journal, type of review (interventional/epidemiologic), continent of authorship, number of authors, year of publication, involvement of a statistician/methodologist, number of syntheses per study and funding.
- Meta-analysis (synthesis) level: magnitude of heterogeneity ($I^2\%$), point estimate along with confidence interval (95% CIs), design of included studies, number of studies within syntheses, type of outcome (binary/continuous), type of analysis (random effects, fixed effects), significance of meta-analysis outcome (significant, non-significant). Standard errors (SE) were calculated for each summary outcome from point estimates and 95%CIs. Heterogeneity ($I^2\%$) was further classified to four non-overlapping categories: 0%, 1- 29%, 30- 59%, 60- 100% based on previous research [17]. Consultation with one author (NP) was performed when uncertainty existed.

Descriptive statistics in relation to the characteristics of the included studies were undertaken at review level and based on publication type (Cochrane, non-Cochrane). Cross-tabulations and associations between level of heterogeneity and synthesis type were applied. Chi-squared and Fisher's exact tests were performed where appropriate. Furthermore, a scatterplot was used to graphically present the relationship between number of studies included within the synthesis and I^2 . Mixed effects univariable and multivariable ordinal logistic regression was used to identify significant predictors for the presence of statistical heterogeneity (0%, 1- 29%, 30- 59%, 60- 100%) both at review level (including review type, involvement of a statistician, number of syntheses) and at synthesis level (including number of studies within syntheses, type of outcome, type of analysis). The random effects parameter was the variable study (ie, SR), which represents the between cluster variability, as more than one meta-analyses within a SR were considered. Non-significant predictors at $p = .10$ in the univariable analysis were excluded from the multivariable model. The presence of significant interaction between type of outcome (binary/ continuous) and type of

analysis (random/ fixed effect) was also assessed. The level of statistical significance was pre-specified at $p < 0.05$. Statistical analyses were performed with STATA version 15.1 software (Stata Corporation, College Station, Tex, USA).

Results

A total of 84 systematic reviews including at least one meta-analysis were initially identified, while 72 remained after applying the eligibility criteria (Figure 1). A quarter of studies were Cochrane reviews ($n=18$, 25%), with 54 being published in orthodontic journals (75%). The majority of the reviews were authored by European investigators (45/72, 63%) and were co-authored by 4-6 researchers (44/72, 61%). Cochrane reviews were more likely to involve a methodologist or statistician than non-Cochrane (18/18, 100%, $p < 0.001$), and also to receive funding (16/18, 89%).

Overall, most reviews included 3 to 5 meta-analyses ($n= 24$, 33%), while there was a considerable increase in the publication of systematic reviews with at least one synthesis in recent years i.e. 2014- 2017 ($n=49$, 68%) (Table 1, Figure 2). The mean review heterogeneity ($I^2\%$, recorded as the average heterogeneity across all syntheses within a review) indicated moderate to considerable I^2 values both within interventional (45/60, 75%) and epidemiologic (10/12, 84%) reviews. A similar finding was confirmed for both Cochrane (11/18, 61%) and non-Cochrane reviews (44/54, 81%) (Table 2).

The total of 72 reviews involved 391 syntheses (Table 3); of these, 125 meta-analyses (32%) reported absence of heterogeneity or heterogeneity explained by chance (ie, $I^2 = 0\%$), while 39 (10%) had I^2 values of 1-29%. Higher levels of heterogeneity were identified within 75 (19%) syntheses ($I^2 = 30-59\%$) with 152 (39%) syntheses having I^2 ranging from 60-100% (Figure 3). Meta-analyses including a larger number of studies were more likely to involve higher levels (ie, $I^2 = 60-$

100%) of heterogeneity (i.e. over 10 studies 21/26, 81%; $p < 0.001$). Likewise, the use of random effects analyses for syntheses of individual trial/study estimates were associated with greater I^2 values (~ 60- 100%: 128/278; 46%) (Table 3).

In the multivariable model, each additional study within individual synthesis was associated with a 20% increase in the odds for substantial/ considerable heterogeneity (ie, $I^2 = 60\text{-}100\%$) compared to the combined lower heterogeneity categories (OR= 1.20; 95%CI: 1.09, 1.31; $p < 0.001$). Use of fixed effect meta-analyses versus random effects was associated with 75% lower odds for substantial/ considerable heterogeneity ($I^2 = 60\text{-}100\%$) compared to the combined lower heterogeneity values categories (OR= 0.25; 95%CI: 0.12, 0.55; $p = 0.001$). The interaction between type of outcome and type of analysis was not significant (Log likelihood ratio test between models with and without interaction: $p = 0.52$, Table 4).

Discussion

The present meta-epidemiologic report revealed that the majority of meta-analyses presented tangible degree of heterogeneity (ie, $\geq 30\%$). It was interesting that meta-analyses of binary outcomes were typically associated with reduced likelihood of considerable heterogeneity and increased reporting of absence of heterogeneity or heterogeneity explained by chance ($I^2 = 0\%$) while the opposite trend was observed for syntheses involving continuous outcomes. However, meta-analyses of binary outcomes were under-represented in our study and no definite conclusions can be drawn. In a similar cross-section of Cochrane and non-Cochrane reviews in medicine, high levels of heterogeneity were found to be twice as common in the assessment of continuous rather than binary outcomes [17], with binary outcomes associated with an increased proportion of heterogeneity explained by chance [19]. A plausible explanation for this may be related to the fact that studies with continuous outcomes may present larger variations in populations, interventions,

outcomes and settings or use of more broad eligibility criteria for sample selection thus contributing to a substantial dissimilarity in the design and other characteristics and subsequent emergence of between-study heterogeneity. An additional assumption might be the fact that studies with continuous outcomes may possibly present greater precision of their point estimates with more tight confidence intervals, thus contributing to greater between study heterogeneity.

A greater number of studies included in a synthesis was found to be associated with increased likelihood of statistical heterogeneity when compared to fewer studies. This may be partially explained by the fact that increasing numbers of studies in a mathematical synthesis, will increase spread and heterogeneity of the sample contributing to overall variation. Notwithstanding this, it should be recognized that I^2 presents low power to detect true statistical heterogeneity when a small number of studies are included in a meta-analysis and interpretation of these findings should be considered with caution [13,20].

This is the first attempt to quantify heterogeneity from meta-analyses and identify potential predictors based on both review level and synthesis level characteristics in dental research. Evaluation was undertaken over a prolonged period with consideration of both Cochrane and non-Cochrane reviews as these are known to differ in relation to quality and reporting characteristics [6]. Consequently, the sample may be regarded as representative of the orthodontic literature evidence base. A plethora of previous research has alluded to similar methodological issues pertaining to a range of dental specialties suggesting that the results from the present investigation are likely to reflect other areas of dental research [21]. In terms of limitations, assessment of heterogeneity within meta-analyses was solely based on reported I^2 . Although this is an intuitive and ubiquitous measure, it does not inform the actual variance of an intervention effect, or the between study dispersion of effects [22,23]. Rather it describes the percentage of variability across

study effects that can be attributed to heterogeneity rather than chance alone⁵. Other measures such as the use of tau-squared (τ^2) are currently available to identify the actual range of effects [24,25]. However, these metrics are not widely available within published reports of meta-analyses, while their interpretation is not as straightforward as the interpretation of I^2 . This meta-epidemiologic study was restricted to review- and syntheses- related characteristics and their association with different levels of heterogeneity. The effect of bias-related reporting characteristics of studies included in meta-analyses such as blinding, random sequence generation, allocation concealment or other reporting related items was not assessed. Although empirical evidence from medical research has shown that between-study heterogeneity may be partially explained by the reporting of characteristics responsible for bias [26], this was beyond the scope of the present work and it may constitute ground for future research.

Conclusions

More than one-third of meta-analyses within the orthodontic literature present either substantial or considerable statistical heterogeneity (based on I^2 values) with increasing heterogeneity observed as the number of studies within the syntheses increased. The consistency of meta-analyses could be improved with more careful consideration of individual study characteristics prior to inclusion in a mathematical synthesis ensuring that the quality of the evidence to inform clinical practice is enhanced.

Acknowledgements

None

Funding

None

References

- [1] C. Mulrow, D. Cook, F. Davidoff, Systematic reviews: critical links in the great chain of evidence. In: Mulrow C, Cook D, eds. *Systematic reviews: Synthesis of best evidence for health care decisions.*, American College of Physicians, Philadelphia, 1998.
- [2] A.M. Glenny, M. Esposito, P. Coulthard, H.V. Worthington, The assessment of systematic reviews in dentistry, *Eur. J. Oral Sci.* 111 (2003) 85–92.
- [3] D. Koletsi, P.S. Fleming, T. Eliades, N. Pandis, The evidence from systematic reviews and meta-analyses published in orthodontic literature. Where do we stand?, *Eur. J. Orthod.* 37 (2015) 603–609. doi:10.1093/ejo/cju087.
- [4] J.P.A. Ioannidis, The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses, *Milbank Q.* 94 (2016) 485–514. doi:10.1111/1468-0009.12210.
- [5] D. Millett, Bias in systematic reviews?, *J. Orthod.* 38 (2011) 158–160. doi:10.1179/14653121141407.
- [6] P.S. Fleming, J. Seehra, A. Polychronopoulou, Z. Fedorowicz, N. Pandis, Cochrane and non-Cochrane systematic reviews in leading orthodontic journals: a quality paradigm?, *Eur. J. Orthod.* 35 (2013) 244–248. doi:10.1093/ejo/cjs016.
- [7] J. Lau, J.P. Ioannidis, C.H. Schmid, Quantitative synthesis in systematic reviews, *Ann. Intern. Med.* 127 (1997) 820–826.
- [8] J.P.T. Higgins, S.G. Thompson, J.J. Deeks, D.G. Altman, Measuring inconsistency in meta-analyses, *BMJ.* 327 (2003) 557–560. doi:10.1136/bmj.327.7414.557.
- [9] G.H. Guyatt, A.D. Oxman, R. Kunz, G.E. Vist, Y. Falck-Ytter, H.J. Schünemann, GRADE Working Group, What is “quality of evidence” and why is it important to clinicians?, *BMJ.* 336 (2008) 995–998. doi:10.1136/bmj.39490.551019.BE.
- [10] G.H. Guyatt, A.D. Oxman, R. Kunz, J. Woodcock, J. Brozek, M. Helfand, P. Alonso-Coello, P. Glasziou, R. Jaeschke, E.A. Akl, S. Norris, G. Vist, P. Dahm, V.K. Shukla, J. Higgins, Y. Falck-Ytter, H.J. Schünemann, GRADE Working Group, GRADE guidelines: 7. Rating the quality of evidence--inconsistency, *J. Clin. Epidemiol.* 64 (2011) 1294–1302. doi:10.1016/j.jclinepi.2011.03.017.
- [11] L.E. Chess, J.J. Gagnier, Applicable or non-applicable: investigations of clinical heterogeneity in systematic reviews, *BMC Med. Res. Methodol.* 16 (2016) 19. doi:10.1186/s12874-016-0121-7.
- [12] J.J. Gagnier, D. Moher, H. Boon, J. Beyene, C. Bombardier, Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature, *BMC Med. Res. Methodol.* 12 (2012) 111. doi:10.1186/1471-2288-12-111.
- [13] CRD, Systematic Reviews. CRD’s guidance for undertaking reviews in healthcare. Centre for Reviews and Dissemination, CRD, University of York, 2008.
- [14] J.P.T. Higgins, S.G. Thompson, Quantifying heterogeneity in a meta-analysis, *Stat. Med.* 21 (2002) 1539–1558. doi:10.1002/sim.1186.
- [15] D.B. Petitti, Approaches to heterogeneity in meta-analysis, *Stat. Med.* 20 (2001) 3625–3633.
- [16] J. Higgins, S. Green, *Cochrane handbook for systematic reviews of interventions*. Version 5.0.0 [updated February 2008]: The Cochrane Collaboration; 2008. Available from: www.cochrane-handbook.org, 2008.
- [17] A.C. Alba, P.E. Alexander, J. Chang, J. MacIsaac, S. DeFry, G.H. Guyatt, High statistical heterogeneity is more frequent in meta-analysis of continuous than binary outcomes, *J. Clin. Epidemiol.* 70 (2016) 129–135. doi:10.1016/j.jclinepi.2015.09.005.
- [18] R. Ryan, Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Group reviews: planning the analysis at protocol stage. <http://cccr.org.cochrane.org>, (2016).
- [19] J. Sterne, M. Bradburn, M. Egger, Meta-analysis in STATA. In: Egger M, Davey Smith G, Altman DG, editors. *Systematic reviews in health care: meta-analysis in context*, 2nd ed., BMJ Publications, London, 2001.
- [20] T.B. Huedo-Medina, J. Sánchez-Meca, F. Marín-Martínez, J. Botella, Assessing heterogeneity in meta-analysis: Q statistic or I² index?, *Psychol. Methods.* 11 (2006) 193–206. doi:10.1037/1082-989X.11.2.193.

- [21] P.S. Fleming, D. Koletsi, K. O'Brien, A. Tschlaki, N. Pandis, Are dental researchers asking patient-important questions? A scoping review, *J. Dent.* 49 (2016) 9–13. doi:10.1016/j.jdent.2016.04.002.
- [22] J.P.T. Higgins, Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified, *Int. J. Epidemiol.* 37 (2008) 1158–1160. doi:10.1093/ije/dyn204.
- [23] G. Rücker, G. Schwarzer, J.R. Carpenter, M. Schumacher, Undue reliance on I^2 in assessing heterogeneity may mislead, *BMC Med. Res. Methodol.* 8 (2008) 79. doi:10.1186/1471-2288-8-79.
- [24] M. Borenstein, J.P.T. Higgins, L.V. Hedges, H.R. Rothstein, Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity, *Res. Synth. Methods.* 8 (2017) 5–18. doi:10.1002/jrsm.1230.
- [25] R.M. Turner, J. Davey, M.J. Clarke, S.G. Thompson, J.P. Higgins, Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews, *Int. J. Epidemiol.* 41 (2012) 818–827. doi:10.1093/ije/dys041.
- [26] K.M. Rhodes, R.M. Turner, J. Savović, H.E. Jones, D. Mawdsley, J.P.T. Higgins, Between-trial heterogeneity in meta-analyses may be partially explained by reported design characteristics, *J. Clin. Epidemiol.* 95 (2018) 45–54. doi:10.1016/j.jclinepi.2017.11.025.

Legends for illustrations

Figure 1. Flow diagram of systematic review selection

Figure 2. Breakdown of orthodontic meta-analyses (%) annually from 2000 to 2017.

Figure 3. Scatterplot of number of studies within each synthesis and magnitude of statistical heterogeneity $I^2\%$.

Table 1. Distribution of study characteristics by journal type (n=72).

	Non-Cochrane	Cochrane	Total	p-value
	N (%)	N (%)		
Continent				0.10*
<i>America</i>	6 (85)	1 (15)	7 (100)	
<i>Europe</i>	30 (67)	15 (33)	45 (100)	
<i>Asio/other</i>	18 (90)	2 (10)	20 (100)	
No. of authors				0.26*
<i>1-3</i>	16 (89)	2 (11)	18 (100)	
<i>4-6</i>	30 (68)	14 (32)	44 (100)	
<i>Over 6</i>	8 (80)	2 (10)	10 (100)	
Statistician Involved				<0.001*
<i>No</i>	51 (100)	0 (0)	51 (100)	
<i>Yes</i>	3 (14)	18 (86)	21 (100)	
Year of Publication				0.23*
<i>2000-2010</i>	10 (63)	6 (37)	16 (100)	
<i>2011-2013</i>	15 (79)	4 (21)	19 (100)	
<i>2014-2017</i>	41 (84)	8 (16)	49 (100)	
Number of syntheses per review				0.72*
<i>1</i>	11 (73)	4 (27)	15 (100)	
<i>2</i>	9 (64)	5 (36)	14 (100)	
<i>3-5</i>	19 (79)	5 (21)	24 (100)	
<i>6-10</i>	8 (89)	1 (11)	9 (100)	
<i>Over 10</i>	7 (70)	3 (30)	10 (100)	
Funding				<0.001*
<i>No</i>	45 (96)	2 (4)	47 (100)	
<i>Yes</i>	9 (36)	16 (64)	25 (100)	
Total	54 (75)	18 (25)	72 (100)	

*Fisher's exact test

Table 2. Distribution of levels of mean heterogeneity based on study type and journal (n=72).

	Mean Heterogeneity I ² %				Total	p-value
	0 N (%)	1-29 N (%)	30-59 N (%)	60-100 N (%)	N (%)	
Type of study						0.17*
<i>Interventional</i>	5 (8)	10 (17)	33 (55)	12 (20)	60 (100)	
<i>Epidemiologic</i>	1 (8)	1 (8)	4 (34)	6 (50)	12 (100)	
Journal						0.19*
<i>Cochrane</i>	2 (11)	5 (28)	6 (33)	5 (28)	18 (100)	
<i>Non-Cochrane</i>	4 (8)	6 (11)	31 (57)	13 (24)	54 (100)	
Total	6 (8)	11 (15)	37 (52)	18 (25)	72 (100)	

* Fisher's exact test

Table 3. Level of heterogeneity according to syntheses variable characteristics (n=391).

	Heterogeneity I ² %				Total	p-value
	0 N (%)	1-29 N (%)	30-59 N (%)	60-100 N (%)	N (%)	
Number of studies within syntheses						<0.001
2	72 (50)	10 (7)	21 (15)	41 (28)	144 (100)	
3-5	44 (25)	23 (13)	37 (21)	71 (41)	175 (100)	
6-10	9 (20)	4 (9)	14 (30)	19 (41)	46 (100)	
Over 10	0 (0)	2 (8)	3 (11)	21 (81)	26 (100)	
Type of study design within synthesis						0.98*
Parallel	114 (31)	38 (10)	72 (20)	142 (39)	366 (100)	
Split-mouth	3 (43)	0 (0)	1 (14)	3 (43)	7 (100)	
Crossover	5 (50)	1 (10)	1 (10)	3 (30)	10 (100)	
Mixed	3 (38)	0 (0)	1 (12)	4 (50)	8 (100)	
Type of analysis						<0.001#
Random effects	68 (25)	26 (9)	56 (20)	128 (46)	278 (100)	
Fixed effect	57 (50)	13 (12)	19 (17)	24 (21)	113 (100)	
Type of outcome						0.33*
Continuous	108 (31)	35 (10)	66 (19)	141 (40)	350 (100)	
Binary	17 (41)	4 (10)	9 (22)	11 (27)	41 (100)	
Significance						<0.001*
No	69 (40)	8 (4)	28 (16)	69 (40)	174 (100)	
Yes	56 (27)	28 (14)	45 (22)	77 (37)	206 (100)	
Non-applicable	0 (0)	3 (27)	2 (18)	6 (55)	11 (100)	
Total	125 (32)	39 (10)	75 (19)	152 (39)	391 (100)	

* Fisher's exact test, # Pearson chi2

Table 4. Mixed effects ordinal logistic regression for the effects of cluster level (review type, type of study, statistician involvement, no. of syntheses per study, funding) and individual level (no. of studies within syntheses, type of outcome, type of analysis, significance of outcome) variables on statistical heterogeneity (heterogeneity categorized as: 0%, 1- 29%, 30- 59%, 60- 100%).

	Univariable			Multivariable		
	Odds Ratio	95%CI	p-value	Odds Ratio	95%CI	p-value
Cochrane						
<i>No</i>	Reference					
<i>Yes</i>	0.55	0.36, 0.86	0.009	2.81	0.53, 14.91	0.22
Type of study						
<i>Interventional</i>	Reference					
<i>Epidemiologic</i>	1.98	0.99, 3.93	0.05	1.47	0.62, 3.47	0.38
Statistician involvement						
<i>No</i>	Reference					
<i>Yes</i>	0.52	0.35, 0.81	0.003	0.64	0.14, 2.94	0.57
No. of syntheses						
<i>Per unit</i>	0.98	0.96, 1.00	0.11			
Funding						
<i>No</i>	Reference					
<i>Yes</i>	0.86	0.53, 1.39	0.55			
No. of studies within syntheses						
<i>Per unit</i>	1.23	1.13, 1.33	<0.001	1.20	1.09, 1.31	<0.001
Type of outcome						
<i>Continuous</i>	Reference			Reference		
<i>Binary</i>	0.55	0.29, 1.06	0.08	0.51	0.23, 1.12	0.09
Type of analysis						
<i>Random effects</i>	Reference			Reference		

<i>Fixed effect</i>	0.23	0.13, 0.43	<0.001	0.25	0.12, 0.55	0.001
Type of outcome x type of analysis interaction						0.52*
Significance of outcome						
<i>No</i>	Reference					
<i>Yes</i>	1.12	0.75, 1.67	0.58			

* Log likelihood ratio test between models with and without interaction

Figure 1.

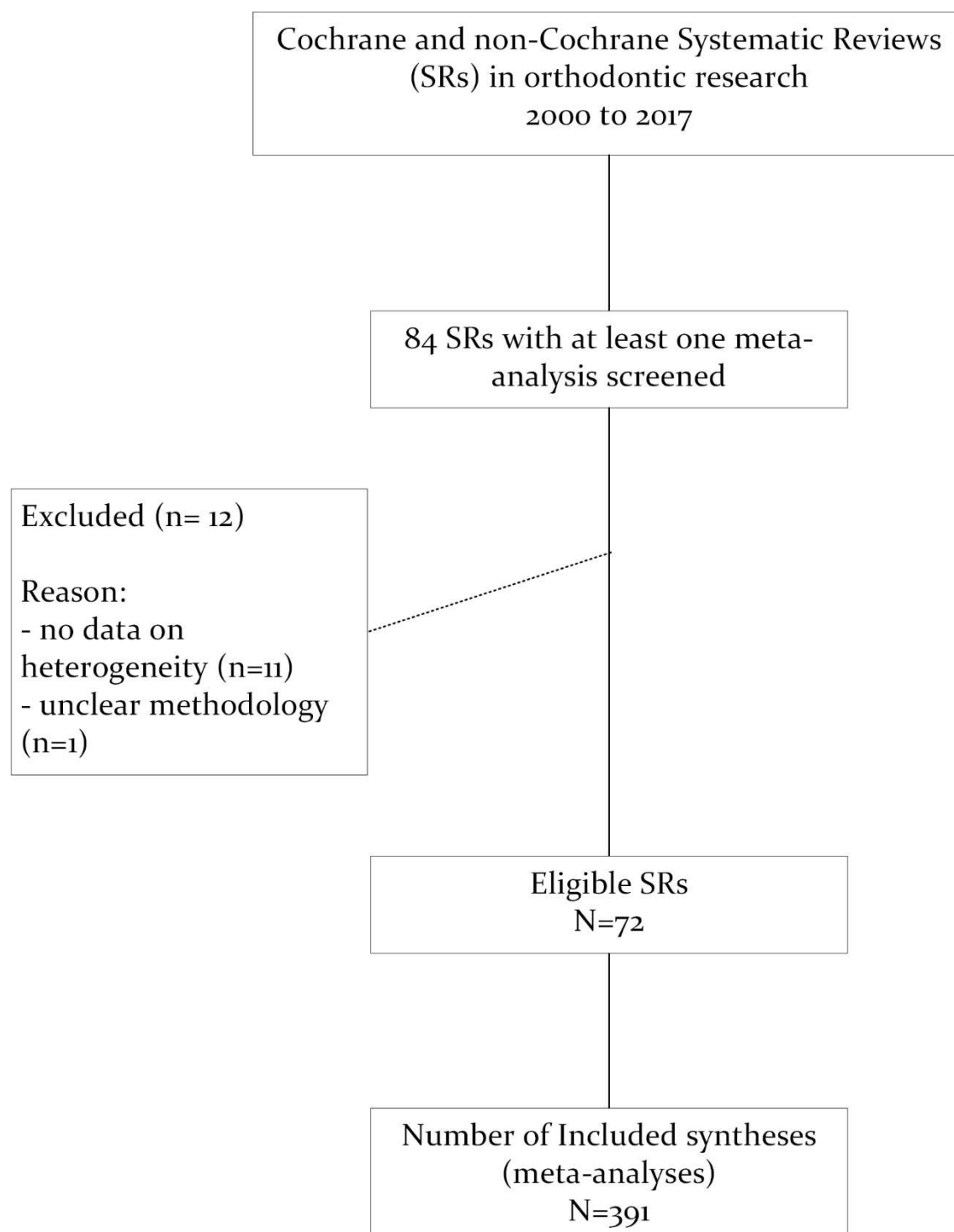


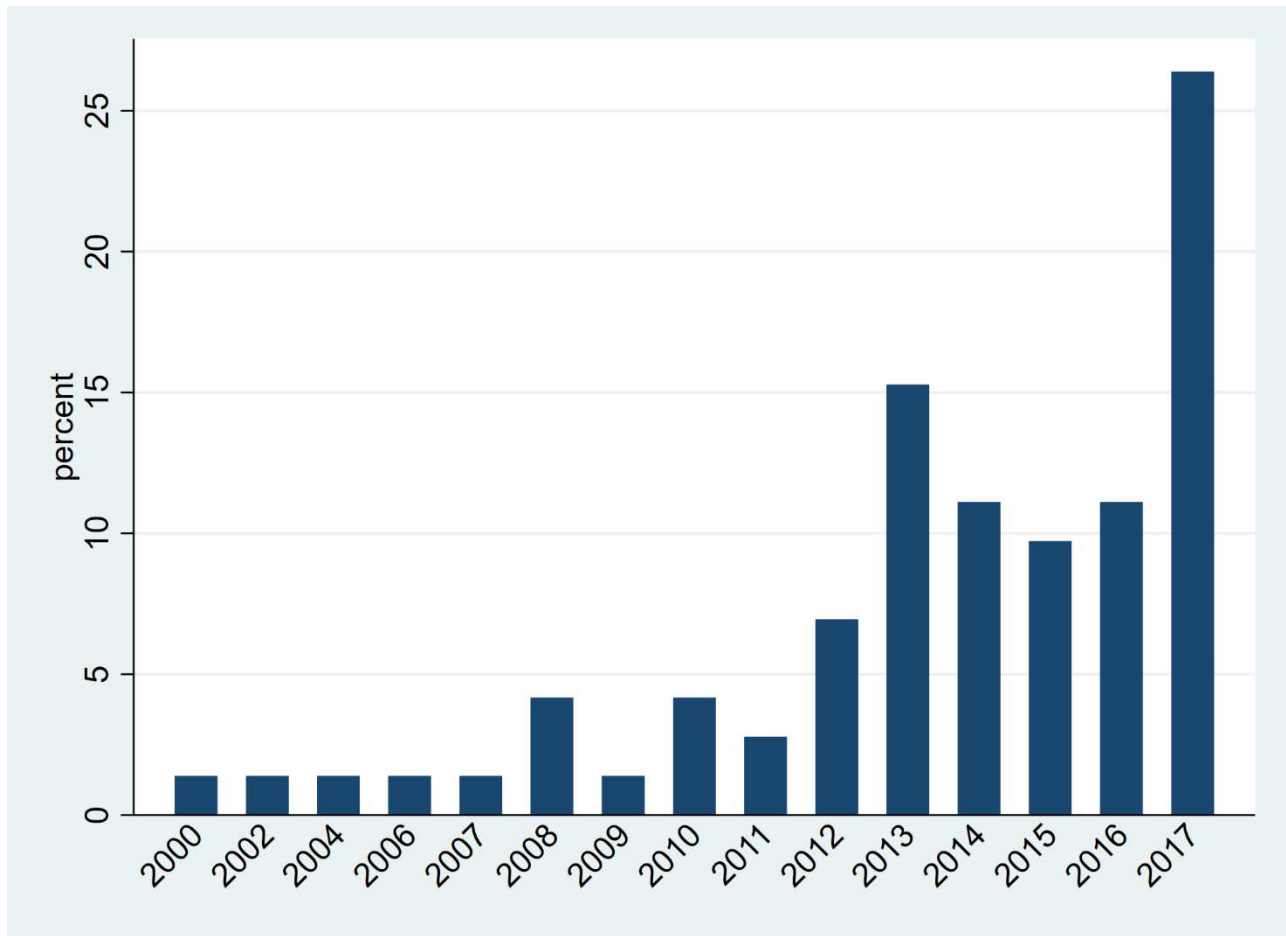
Figure 2.

Figure 3.

